基于图像级标签及超像素块的弱监督显著性检测

谭台哲 [1,2]，轩康西 [1+]，曾群生 [1]

(1. 广东工业大学 计算机学院，广州 510006; 2. 河源广工大协同创新研究院，广东 河源 517000)

摘 要：针对获得训练数据集代价高昂问题，提出了一种用于图片显著性检测的弱监督新方法，在训练网络模型时仅使用图片级标签。方法分为两个阶段，在第一阶段，根据图片级标签训练分类模型，获得前景推断图；在第二阶段，对原图片进行超像素块处理，并与阶段一得到的前景推断图进行融合，从而细化显著对象边界。算法使用了现有的大型训练集和图像级标签，未使用像素级标签，从而减少了注释的工作量。在四个公共基准数据集上的实验结果表明，性能明显优于无监督的模型，与全监督模型相比也具有一定的优越性。

关键词：深度学习；弱监督；显著性检测；超像素

中图分类号：TP391.41 doi: 10.19734/j.issn.1001-3695.2018.06.0576

Supervised significant detection based on image level labels and superpixel blocks

Tan Taizhe[1, 2], Xuan Kangxi[1†], Zeng Qunsheng[1]

(1. *College of Computer, Guangdong University of Technology, Guangzhou* 510006, *China*; 2. *Heyuan Guanggong Collaborative Innovation Research Institute, Heyuan Guangdong* 517000, *China*)

Abstract: Aiming at the high cost of obtaining the training data set, proposing a new weak supervision method for image saliency detection. Only using the picture-level label when training the network model. Dividing the method into two stages. In the first stage, training the classification model according to the picture-level label to obtain the foreground inference graph. In the second stage, processing the original image by super-pixel block and merged with the foreground inference graph obtained in phase one, thus refining significant object boundaries. The algorithm uses existing large training sets and image-level tags, eliminating the use of pixel-level tags, which reduces the amount of annotation work. The experimental results on the four common benchmark datasets show that the performance is significantly better than the unsupervised model, and it has certain advantages compared with the full-supervised model.

Key words: deep learning; weak supervision; significance detection; superpixel

0 引言

图像效果是神经科学和心理学中一个重要的基础研究问题，是用于研究人类视觉系统从复杂场景中选择感兴趣区域的机制。人类有能力准确快速地发现感兴趣的对象（或区域），这就是所谓的焦点或显著的情景。在突出刺激的驱动下，注意力被认为是部分自由的、自下而上的、无记忆的。注意力也可以由相对缓慢的、自上而下的记忆依赖机制来指导。比如，当人们看人脸时，所熟悉的面孔可能会先一步引起人们的注意。可靠的视觉显著性估计使得即使在没有先验知识的情况下也可以对图像进行适当的处理。因此，视觉显著性是许多计算机视觉任务的重要步骤。

近几年，在计算机视觉领域取得重大进展的卷积神经网络（CNN）引起了人们的广泛关注，兴起了使用精确的像素级注释样本进行图片显著性检测的浪潮[1~3]。与无监督方法[4,5]相比，基于全监督机制学习的 DNN 更有效地捕获语义上突出的前景区域，在复杂场景下产生准确的结果。 但是，鉴于 DNN 的数据饥饿性质，其卓越的性能也严重依赖于大量数据集与像素级注解进行训练。然而，注释工作非常繁琐，精确注释的训练集仍然稀少且昂贵。

为了减轻大规模像素级注释的需要，本文研究了图像级标签的弱监督方法来训练显著性检测器。图像级标签表示图像中存在的对象类别，并且比像素方面的注释更容易收集。同时，图像级标签提供了很可能是显著前景的图像中主要对象的类别信息。此外，最近的工作[6,7]已经提出，只有图像级别标签训练的 DNN 也提供了对象位置信息。因此，这种仅利用图像级标签的来训练 DNN 来检测突出物体的弱监督方法是行之有效的。

尽管使用 DNN 能得到图像可视化后能明显地提取出前景目标，但是在边缘处仍然是模糊的，这是因为边界周围的像素集中在相似的感受野，所以需要对显著图做边界细化处理。因此，本文的实验分为两个阶段：使用图像级标签的预训练和结合超像素块的边界细化。

在第一阶段，鉴于池化层会损失大量的细节信息，利用图像级标签预训练了一个全卷积网络（FCN），通过改变卷积核在图片的滑动步长来代替池化层，从而获得多尺度的显著特征。第二阶段，受文献[8]的启发，提出了一种卷积特征-超像素边界联合细化的全新方法。首先整合第一阶段得到的特征图到其特征边界（FB），然后对原图进行超像素处理获得超像素边界（SPB）。根据 SPD 调整 FB，从而达到细化边界的目的。

本文的贡献包括两个方面。首先，为弱监督显著性检测

收稿日期：2018-06-14；**修回日期：**2018-08-15

作者简介：谭台哲（1970-），男，山东莱阳人，副教授，博士，主要研究方向为机器学习与大数据处理、图像处理与计算机视觉；轩康西（1993-），男（通信作者），硕士，主要研究方向为深度学习、图像处理（1056808552@qq.com）；曾群生（1993-），男，硕士，主要研究方向为深度学习、图像处理.

提供了一种新的方向，只使用现有的大量图像级标签，从而极大减少了注释的工作量；其次，提出了一种新颖的细化边界方法，更好地利用了原图的细节信息，从而弥补了卷积神经网络边界模糊的不足，进一步提高了检测的准确率。

1　相关工作

很多传统图像处理算法，像 CRFs[9]、随机森林[10~12]，SVM[13]等已成功应用在显著性检测问题上。这些现有的方法旨在通过在图像上找到图结构来捕获图像的上下文信息以及使用分类器标记不同的实体，如超像素等[14~16]。Jiang 等人[10]将显著性检测当做一个回归问题，在对图像进行超像素处理后，使用监督学习的方法将特定区域的特征向量映射到显著分数上，在训练结束后，再整合成一张显著图。中 Li 等人[18]通过训练一个 SVM 来检测图像中的显著物体，与此同时使用了超边缘分块以及多尺度方法进行后续处理。

但是，基于 DNN 的方法却证明了其在显著性检测方面的巨大优势，其中，基于 FCN 的显著性检测方法[3,19]在准确性和速度方面更具有竞争力的表现。Wang 等人[21]通过整合局部估计和全局搜索来预测显著性图。文献[22]中提出了一个双阶段深度网络，首先生成一个粗略图，然后使用另一个网络逐步对其进行分级细化。可训练这些模型需要大量的像素级注释，这种方法的成本是非常昂贵的。文献[23]作为弱监督检测的代表，提出了前景推断网络 FIN 和迭代条件随机场这一新模型，其性能明显优于无监督算法，甚至优于全监督算法。综上所述，使用 DNN 模型结合图像级标签的弱监督显著性检测方法也是解决图像显著性问题的一个新方向。

2　弱监督显著性检测

用于图像级标签预测的 CNN 通常由一系列卷积层组成，然后是几个全连接层。假设用 I 代表训练图片，l∈{1，2，…，N}代表其对应的类别标签。CNN 就是以 I 作为输入，在一系列计算后得到一个 N 维的分数向量 Y，Y 中最大值的索引即为该图片的类别。同时，训练 CNN 模型时需要最小化损失函数 L 来衡量预测值准确情况。虽然 CNN 模型基于图像级标签进行训练，但最近的实验证明，高层次的卷积层有能力作为检测器捕获并识别目标物体部分。可另一方面，卷积层编码的目标位置信息无法转换成全连接层的编码。

因此，在多标签识别任务中 Jonathan long 等人提出了全卷积网络（FCN）来保存目标的位置信息。给定一张输入图片 H*W，使用的像素级注释作为监督信号，经过模型训练后，输出 N 个通道的 H*W 分数图，每一个通道代表一个类别，这 N 个通道中对应点的值代表图片对应像素点属于此类的可能性。可是，FCN 提取的显著图边缘却非常模糊。所以，在弱监督显著性检测中，本文改最后的输出层为 N*1*1 的分数图，这 N 个值分别代表这张图片属于此类的可能性；然后通过整合高层次的卷积层特征图得到一张前景推断图，接着对推断图做进一步的后续处理。

2.1　前景推断图

在以图像级标签为监督信号的 FCN 模型训练时，其卷积核能捕捉到输入图像的对象区域，每个通道对应对象的一个特征。在显著性检测任务中，本文不关心对象类别，旨在发现所有的显著对象区域。为了获得这样的与类别无关的显著图，可以将同一尺度所有通道的特征图求和，然后再映射到 0-255 的颜色值之间进行可视化[24]。但是，这样做有个缺点：显著对象的部分响应会被其他通道的较高响应区域抑制，因此产生的显著图要么有大量的背景噪声，要么不能均匀地高亮显著区域。

所以，为了解决上述问题，本文在训练 FCN 时添加了一个分支来自动生成前景推断图（FIM）。这一分支也是由一系列的卷积层和一个 sigmoid 层组成。输出的特征图 F 仅有一个通道，数值范围在[0,1]，代表了对应像素点的显著度。总的来说，给定一张图片 X，经过模型计算，分别生成 C（通道数）张特征图 S(n*n)和一张前景推断图 F，代入以下公式：

$$S_k^{'} = S_k \odot F \tag{1}$$

其中：S_k 代表特征图 S 的第 k 个通道，⊙ 代表 S_k 和 F 对应元素相乘，$S_k^{'}$ 代表两者整合后的特征图传递到下一层。这样，通过利用特征图 S 中的每个通道的高响应，不让它们相互抑制，使得 FIM 的生成也有一个不断学习和训练的过程 [25,26]。

同时考虑到使用特定的图像级标签训练的 FCN 很难覆盖到不在训练集中的类别，所以将式（1）的蒙板操作应用于中间提取的特征图，而不是最后一层。因为中间层特征图并不直接和图片的类别相对应，而是提取出特定的结构、纹理等，这些表达特征的方式是通用的，这样，生成的 FIM 能够更好地捕捉到训练时为出现过的新类别，提高了模型的鲁棒性。

图 1 为本文的网络结构。在模型的第一阶段通过训练 FCN（1）~（5）来生成一张前景推断图 FIM。在第二阶段结合超像素块显著度图（6）对 FIM 进行边缘细化生成最终的显著图（7）。

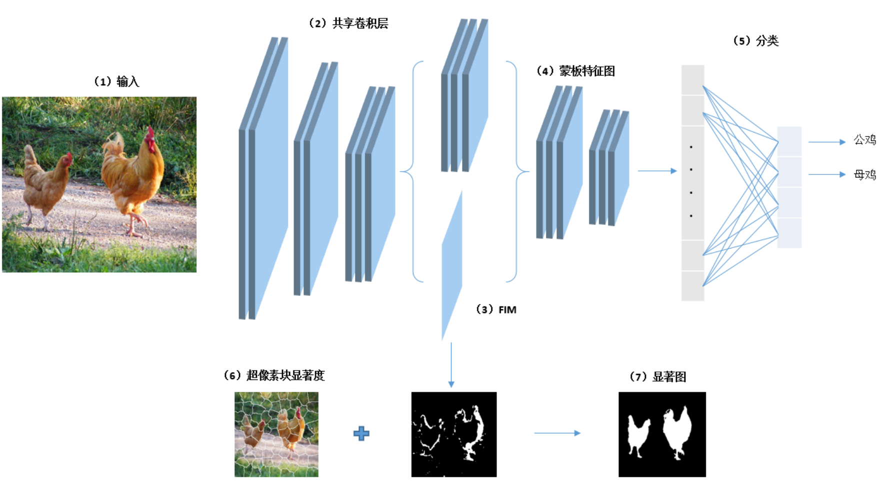


图 1　本文设计的网络结构

Fig. 1　Designed network structure in this paper

2.2　基于图像级标签的预训练

从这一节中正式开始介绍弱监督显著性检测方法的第一阶段。使用 ImageNet 数据集训练网络，这一数据集中包括 1000 个对象类别，每类 1000 张图片。

正如上述讨论，由于 FIM 的生成和和 FCN 的训练有着密切的联系，所以它们可以联合训练并共享卷积特征。具体实现是，在 16 层的 VGG 网络[27]后设计共享网络分支，该 VGG 网络由 13 个卷积层组成，卷积层之间由 ReLU 非线性函数和 4 个最大化池连接。FIM 分支网络由一个卷积层、一个 BN 层[28]和一个 Sigmoid 层计算得到，然后作为蒙板和 FCN 整合得到新的特征图，最后通过全连接层生成 1000 个对象类别得分向量，并使用 Softmax 函数将得分向量转换成类别概率。

对于每个图像生成的显著图，较大的值意味着该像素更可能属于前景。通过大量的观察可以推断出前景像素和语义对象之间存在显式关联。由于每个简单图像都附带有语义标签，因此可以容易地推断出可以为前景候选像素分配相应的图像级标签。然后，提出了一种多标签交叉熵损失函数来训练显著图监督下的分割网络。

给定一个包含 N 个训练样本的训练集 $\{X_i, L_i\}_{i=1}^{N}$，本文采用最小化下面的损失函数[23]来达到使模型收敛的方法：

$$\min_{\theta} -\frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{k\in l_i}\log\big(p(k|\boldsymbol{X}_i;\boldsymbol{\theta})\big) + \sum_{k\notin l_i}\log(1-p(k\mid \boldsymbol{X}_i;\theta)) - \lambda\|\boldsymbol{f}(\boldsymbol{X}_i;\boldsymbol{\theta})\|_1\Big] + \eta\theta_2^2, \tag{2}$$

其中：θ 代表网络参数，第一和第二个参数使保证预测准确率的交叉熵损失，第三个参数是 FIM 的 L1 正则化，最后一个参数是网络衰减参数，根据经验，λ 和 η 分别设置为 7e-5 和 5e-5。共享层的权重参数使用预训练的 VGG 模型初始化[27]，其他层的权重使用文献[29]的方法进行随机初始化。所有输入图像都归一化到 224*224 的固定分辨率，而 FIM 的分辨率为 56*56，然后通过双线性内插法放大到 224*224。为了使上述损失函数快速收敛，本文采用随机梯度下降（SGD）方法。

2.3　基于超像素块的边缘检测

在 DNN 训练完成后可以得到 FIM，但正如前面说到的，FIM 的边缘比较模糊，所以，在这一部分将对 FIM 做进一步的后续处理来细化其轮廓。

对比度是评估人类视觉的重要参数，由于显著物体和周围环境的对比度是不同的，且人类的视觉细胞对图像边缘更加敏感，所以通过对比度计算确定图像的边缘，进而将图片分割成超像素块。传统的图像处理方法根据图片的三种属性对图片进行处理：颜色、纹理和形状[8,30]。这些技术已经成功应用在各个方面。但是，这些属性无法提供对图像的高度理解，因为人类通常不会单独从颜色、纹理或者形状去理解图像，而是基于这三个属性特征背后的相互联系，也就是说，一张图片中目标物体的显著程度取决于其与周围环境的独特性。

根据笔者的观察，显著对象有三个明显特征，由此可以计算出显著对象的形状特征：a) 显著对象总是和其周围的环境明显不同；b) 显著对象几乎都位于图像的中心附近；c) 显著对象的边界都是完美闭合的。

第一个特征基于自下而上的视觉刺激，人们[9, 31~33]对此已经做了大量的研究；第二个是位置优先特征，人们的注意力总是先关注图片的中心位置，然后再向四周发散[3~37]；第三个特征由[18,31]提出，显著目标通常是聚在一起的，而不是分散在图片各处，且人们在观察物体时，也是观察物体的边缘，然后大脑再将这些边缘进行组合，形成物体。

本文的目标是将图片分割成一个个封闭的轮廓，其边界包含了图像中显著目标对象的边界。首先，通过边缘检测对图像进行分割，进而生成超像素块，然后根据超像素图像计算对应的显著度。文献[9,31]将原始图像缩放为较小的尺寸来减少计算量。而本文的方法得到的图像中的超像素块数量远小于像素数量，因此在减少计算量的同时，也能生成全分辨率的边缘图。

在图像分割问题上分两方面考虑：颜色距离和空间距离。

本文利用文献[40]的方法将图片分成若干个区域，然后根据颜色对比度评估每一个区域基于颜色值的显著度。根据特征一、二，如果一个区域（超像素块）与其周围的上下文信息明显不同，同时，其与图片中心的点位置较近，则此区域（超像素块）显著的可能性较大。同时，对于属于同一个类别的超像素块来说，一方面，无论空间上距离多远，它们的相似度总是很高的；另一方面，距离过远的相似超像素块无法明显相互提高对方的显著度。

综上所述，对于一张含有 N 个超像素块的图像来说，先计算中心超像素块的显著性：

$$\begin{cases} S_{center} = 1 - exp\left[-\frac{1}{N-1}\sum_{n=1}^{N-1} d(p_i, q_n)\right] \\ d(p_i, q_n) = w_c d_{color}(p_i, q_n) w_p d_{position}(p_i, q_n), \end{cases} \tag{3}$$

其中：d_{color} 和 $d_{position}$ 分别代表超像素块 p_i、q_n 的颜色距离和空间距离，w_c 和 w_p 是两个超参数，分别用来表示颜色距离和空间距离的强度大小。$d(p_i, q_n)$ 代表超像素块 p_i、q_n 的相异度。

对于其他非中心超像素块，根据特征 2，本文额外引入了中心超像素块显著度的影响：

$$S = 1 - exp\left\{-\frac{1}{N-1}\sum_{n=1}^{N-1} d(p_i, q_n)\right\} + d(p_i, q_c) S_{center}, \tag{4}$$

q_c 在这里指处于中心超像素块。通过上述公式可以得到图像中每一个像素块的显著度。

2.4　FIM 和超像素块联合边界细化

通过 2.2 节的预训练后，生成的 FIM 已经捕捉到了前景区域。正如人们所知道的，FIM 中包含了大量的边缘信息，通过设置一个阈值生成 FIM 的二值化图，如图 1 所示。在对 FIM 进行二值化时，每张图片使用的阈值都是不一样的，即设置的阈值不是一个固定值，而是先对图像作降噪处理，以去掉 FIM 中高响应的噪声点，然后通过计算图像的直方图，再次将图像中出现频率低的像素点用相似的像素点代替[30]，最后取最大和最小像素值的一个中间值作为阈值。但是，可以看到，图像中物体的边界仍然是不连续的，所以需要对这些边界做连通处理。

根据显著对象的第三个特征：显著对象的边界都是完美闭合的，首先通过计算连通区大小找到 FIM 的最长边缘，取其一个端点作为起点，根据梯度优先的规则搜索附近的其他边缘，如图二所示，端点 A 的延伸趋势是向下，所以它会优先向下搜索未连通的边缘。当找到需要连通的端点 B 后，A 点就会根据 2.3 节得到的显著分数图向 B 点延伸。

图 2 为 FIM 边缘细化过程。图中(a)为超像素块显著图，

(b)为二值化后的前景推断图 FIM，(c)为边缘细化后的 FIM，其中红色的线是结合超像素块显著度延伸的边缘线，(d)是最终的显著性分割结果。

3　实验

在超参数的设计上，本文提出的方法基于 TensorFlow 实现，权重衰减取 0.001，冲量取 0.9。FIM 二值化时若像素值的频率小于 10 则对作相似点代替处理，在得到即对 FIM 中的像素点若其显著度大于阈值则将其设置为 255，反之则置为 0。

在实验方面，本文方法和 MBS[41]，wCtr[47]、MR[48]、BSCA[42]、WSS[5]、RFCN[3]、DCL[22]、DS[49]、MC[1]这九个模型比较，由于显著性检测是一个新的视觉问题，发布的数据集也十分有限，所以本文选取了 4 个公共数据集进行测试：

1) SED[45]　包含 100 张图片，每张图片包含一个类别；

2) ECSSD[46]　包含 1000 张结构复杂的图片，每张图片中有多个类别；

3) MSRA-B[9]　包含 5000 张图片，200 多个类别；

4) PASAL-S[12]　从 PASCAL VOC 数据集中精心挑选的复杂环境下的 850 张图片

在比较时，本文引入了 F_β 作为效果指标，如图 3 所示。对最终得到的显著图进行二值化，并与像素级真值注释进行对比，可以得到一组正确值和召回值。每一个数据集的 F_β 是从所有图像的平均精度和召回值得到，F_β 的定义为

$$F_\beta=\frac{(1+\beta^2)Precision\times Recall}{\beta^2 Precision+Recall},\tag{5}$$

其中，β^2 值为 0.3。通过与现今最好的方法进行比较以证明本文方法的有效性。

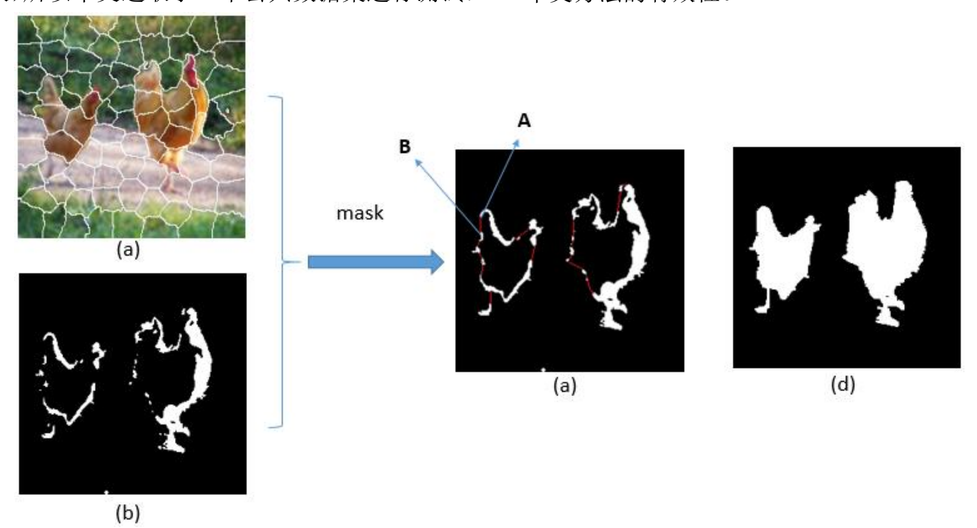
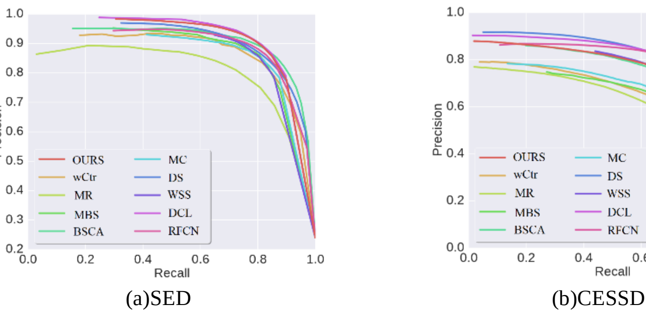
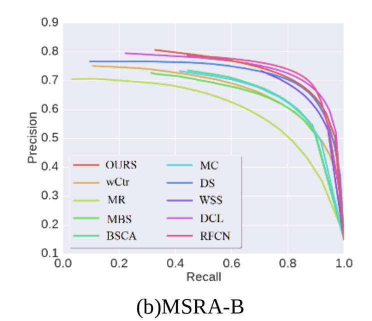


图 2　FIM 边缘细化过程

Fig. 2　FIM edge thinning process





(a)SED　(b)CESSD　(b)MSRA-B

图 3　precision-recall 曲线

Fig. 3　precision-recall curve

表 1　算法测试比较

Table 1　Test comparison of algorithms

	Unsupervised				Weakly		Fully			
	MBS	wCtr	MR	BSCA	WSS	This	RFCN	DCL	DS	MC
SED	0.776	0.786	0.782	0.756	0.838	**0.841**	0.813	0.825	0.794	0.817
ECSSD	0.673	0.676	0.69	0.705	0.823	0.817	**0.834**	0.829	0.826	0.796
MSRA-B	0.726	0.731	0.729	0.735	0.783	0.792	**0.811**	0.802	0.787	0.763
PASAL-S	0.604	0.597	0.583	0.597	0.72	0.733	**0.747**	0.71	0.655	0.687

本文选择了当前最优的四个无监督算法、一个弱监督算法和四个全监督算法作为比较，结果如表 1 所示。并引入 F_β 来衡量各方法的检测效果，其中黑体部分是指当前数据集上最好结果。

根据表 1 可以看出，由于本文的方法在提取特征时是有目的和一定的监督信号进行不断调优的，所以始终优于无监督方法。同时，由于没有使用获取代价高昂的像素级注释，但最终的实验效果却和全监督方法相比也有一定的优势。另外，大多数全监督显著性检测数据集包含的图片虽然很多，但类别却不足 300，而本文的方法基于 ImgNet 进行训练，提取到的类别特征相应也很多，所以本文的方法具有更好的鲁棒性。

4　结束语

本文提出了一种基于图像级标签的弱监督显著性检测方法，此方法分为两个阶段，在第一阶段，在 FCN 的基础上添加了新颖的一层，通过学习预测图像级标签来生成一张前景推断图 FIM。在第二阶段，根据显著对象的三个特征并结合上下文信息对输入图片进行超像素处理，计算每一个超像素

块的显著度，然后以超像素块的显著度为依据对 FIM 边缘进行细化处理。通过在基准数据集上的评估验证了本文方法的有效性。

参考文献：

[1] Zhao Rui, Ouyang Wanli, Li Hongsheng, *et al.* Saliency detection by multi-context deep learning [C]// Computer Vision and Pattern Recognition. IEEE, 2015: 1265-1274.

[2] Li Guanbin, Yu Yizhou. Visual saliency based on multiscale deep features [C]// Computer Vision and Pattern Recognition. IEEE, 2015: 5455-5463.

[3] Wang Linzhao, Wang Lijun, Lu Huchuan, *et al.* Saliency detection with recurrent fully convolutional networks [C]//Proc of European Conference on Computer Vision. Berlin:Springer, 2016: 825-841.

[4] Zhang Jianming, Stan Sclaroff, Lin Zhe, *et al.* Minimum barrier salient object detection at 80 fps [C]//Proc of IEEE International Conference on Computer Vision. Washington DC:IEEE Computer Society, 2015: 1404-1412.

[5] Kong Yuqiu, Wang Lijun, Liu Xiuping, *et al.* Pattern mining saliency [M]// Computer Vision. Springer International Publishing, 2016: 583-598.

[6] Long Jonathan, Zhang Ning, Trevor Darrell. Do convnets learn correspondence? [C]//Advances in Neural Information Processing Systems.2014: 1601-1609.

[7] Zhou Bolei, Aditya Khosla, Antonio Lapedriza, *et al.* Learning deep features for discriminative localization [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition.Washington DC: IEEE Computer Society, 2015: 2921-2929.

[8] Jiang Huaizi, Wang Jingdong, Yuan Zejian, *et al.* Automatic salient object segmentation based on context and shape prior [C]//Proc of British Machine Vision Conference. 2011.

[9] Liu Tie, Sun Jian, Zheng Nanning, *et al.* Learning to detect a salient object. [J]. IEEE Trans on Pattern Anal Mach Intell, 2011, 33(2): 353-367.

[10] Jiang Huaizu, Wang Jingdong, Yuan Zejian, *et al.* Salient object detection: a discriminative regional feature integration approach [J]. International Journal of Computer Vision, 2017, 123(2): 251-268.

[11] JiwhanKim, Han Dongyoon, Tai Yu-Wing, *et al.* Salient region detection via high-dimensional color transform [J]. IEEE Trans on Image Processing, 2015, 25(1): 9-23.

[12] Li Yin, Hou Xiaodi, Christof Koch, *et al.* The secrets of salient object segmentation [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 280-287.

[13] Lu Song, Vijay Mahadevan, Nuno Vasconcelos. Learning optimal seeds for diffusion-based salient object detection [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 2790-2797.

[14] Joseph Tighe, Marc Niethammer, Svetlana Lazebnik. Scene parsing with object instances and occlusion ordering [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3748-3755.

[15] Guo Ruiqi, Derek Hoiem. Labeling complete surfaces in scene undierstanding [J]. International Journal of Computer Vision, 2015, 112 (2): 172-187.

[16] Nasim Souly, Mubarak Shah. Scene labeling using sparse precision matrix [C]// Computer Vision and Pattern Recognition. IEEE, 2016: 3650-3658.

[17] Jiang Huaizu, Wang Jingdong, Yuan Zejian, *et al.* Salient object detection: a discriminative regional feature integration approach [J]. International Journal of Computer Vision, 2017, 123 (2): 251-268.

[18] Li Xi, Li Yao, Shen Chunhua, *et al.* Contextual hypergraph modeling for salient object detection [C]//Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2014: 3328-3335.

[19] Jason Kuen, Wang Zhehua, Wang Gang. Recurrent attentional networks for saliency detection [C]// Computer Vision and Pattern Recognition. IEEE, 2016: 3668-3677.

[20] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Distance transforms of sampled functions [J]. Theory of Computing, 2004, 8(19): 415-428.

[21] Wang Lijun, Lu Huchuan, Ruan Xiang, *et al.* Deep networks for saliency detection via local estimation and global search [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:IEEE Computer Society, 2015: 3183-3192.

[22] Li Guanbin, Yu Yizhou. Deep Contrast learning for salient object detection [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:IEEE Computer Society, 2016: 478-487.

[23] Wang Lijun, Lu Huchuan, Wang Yifan, *et al.* Learning to Detect salient objects with image-level supervision [C]// Computer Vision and Pattern Recognition. IEEE, 2017: 3796-3805.

[24] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//Proc of European Conference on Computer Vision. Berlin: Springer, 2014: 818-833..

[25] Kelvin Xu, Jimmy Ba, Ryan Kiros, *et al.* Show, attend and tell: neural image caption generation with visual attention [J]. Computer Science, 2015: 2048-2057.

[26] Dai Jifeng, He Kaiming, Sun Jian. Convolutional feature masking for joint object and stuff segmentation[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 3992-4000.

[27] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014.

[28] Sergey Ioffe, Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. 2015: 448-456.

[29] He Kaiming, Zhang Xiangyu, Ren Shaoqig, *et al.* Delving deep into rectifiers: surpassing human-level performance on imagenet classification [C]//Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2015: 1026-1034.

[30] Cheng Mingming, Zhang Guoxin, N J Mitra, *et al.* Global contrast based salient region detection [C]// Computer Vision and Pattern Recognition. IEEE, 2011: 409-416.

[31] Stas Goferman, Lihi Zelnik-manor, Ayellet Tal. Context-Aware Saliency Detection [C]// Computer Vision and Pattern Recognition. IEEE, 2010: 2376-2383.

[32] Hou Xiaodi, Zhang Liqing. Saliency detection: a spectral residual approach [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:IEEE Computer Society, 2007: 1-8.

[33] Laurent Itti, Christof Koch, Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2002, 20(11): 1254-1259.

[34] Vladimir Kolmogorov, Ramin Zabih. What energy functions can be minimized via graph cuts? [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2004, 26(2): 147-159.

[35] Bhattacharya Subhabrata, Sukthankar Rahul, Shah Mubarak. A framework for photo-quality assessment and enhancement based on visual aesthetics [C]//Proc of ACM International Conference on Multimedia. New York:ACM Press, 2010: 271-280.

[36] Ritendra Datta, Dhiraj Joshi, Li Jia, *et al.* Studying aesthetics in photographic images using a computational approach [C]//Proc of European Conference on Computer Vision. Springer-Verlag, 2006: 288-301.

[37] Luo Yiwen, Tang Xiaoou. Photo and video quality evaluation: focusing on the subject [C]//Proc of European Conference on Computer Vision. Springer-Verlag, 2008: 386-399.

[38] Joachim S. Stahl, Wang Song. Edge grouping combining boundary and region information [J]. IEEE Trans on Image Processing,2007, 16 (10): 2590-2606.

[39] Sara Vicente, Vladimir Kolmogorov, Carsten Rother. Graph cut based image segmentation with connectivity priors [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.

[40] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59 (2): 167-181.

[41] Zhang Zhiqi, Cao Yu, Dhaval Salvi, *et al.* Free-shape subwindow search for object localization [C]// Computer Vision and Pattern Recognition. IEEE, 2010: 1086-1093.

[42] Qin Yao, Lu Huchuan, Xu Yiqun, *et al.* Saliency detection via cellular automata [C]// Computer Vision and Pattern Recognition. IEEE, 2015: 110-119.

[43] Zhang Dingwen, Meng Deyu, Han Junwei. Co-saliency detection via a self-paced multiple-instance learning framework [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39 (5): 865-878.

[44] Philipp Krähenbühl, Vladlen Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials [EB/OL]. 2012. https://arxiv.org/pdf/1210.5644.pdf.

[45] Sharon Alpert, Meirav Galun, Ronen Basri, *et al.* Image Segmentation by probabilistic bottom-up aggregation and cue integration [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. IEEE Xplore, 2007: 1-8.

[46] Yan Qiong, Xu Li, Shi Jiaya, *et al.* Hierarchical saliency detection [C]// Computer Vision and Pattern Recognition. IEEE, 2013: 1155-1162.

[47] Zhu Wangjiang, Liang Shuang, Wei Yichen, *et al.* Saliency optimization from robust background detection [C]// Computer Vision and Pattern Recognition. IEEE, 2014: 2814-2821.

[48] Yang Chuan, Zhang Lihe, Lu Huchuan, *et al.* Saliency detection via graph-based manifold ranking [C]// Computer Vision and Pattern Recognition. IEEE, 2013: 3166-3173.

[49] Li Xi, Zhao Liming, Wei Lina, *et al.* Deepsaliency: multi-task deep neural network model for salient object detection [J]. IEEE Trans on Image Processing, 2016, 25(8): 3919.